

Do Expressive Geographic Queries Lead to Improvement in Retrieval Effectiveness?

Damien Palacio^{1,3}, Christian Sallaberry¹, Guillaume Cabanac², Gilles Hubert²,
and Mauro Gaio¹

¹ Université de Pau et des Pays de l'Adour, LIUPPA ÉA 3000
Avenue de l'Université, BP 1155, F-64013 Pau cedex, France

² Université de Toulouse, IRIT UMR 5505 CNRS
118 route de Narbonne, F-31062 Toulouse cedex 9, France

³ GIS, Geographisches Institut, Universität Zürich,
Winterthurerstrasse 190, 8057 Zürich, Switzerland

damien.palacio@geo.uzh.ch
{christian.sallaberry, mauro.gαιο}@univ-pau.fr
{guillaume.cabanac, gilles.hubert}@irit.fr

Abstract. In an information retrieval (IR) context, users usually issue queries with few terms and no operators (e.g., and, or, +). However, most of users' information needs involve more expressiveness (e.g., *'Potato famine in Ireland, but not in Cork'*). Our work deals with this category of queries that may be processed by geographic IR (GIR) systems to parse digital libraries according to spatial, temporal and topical criteria. We propose a GIR framework that supports expressive queries and aggregates results of a multi-criteria search. We also conduct experiments to verify that this approach improves the effectiveness of such search engines (improvement of 27% for topical criteria only, and of 54% for spatial and temporal criteria).

1 Introduction

Several studies [26,27,31] on popular search engines, such as Altavista, Excite, Google, MSN Search or Yahoo!, showed that the users of web-based search engines typically issue queries with few terms (between 2 and 3) and rarely use operators (up to 20%). In the same time, Lucas and Topi [17] analyzed the effects of term and operator usage on the relevance of search results. They showed that queries from experts involve more terms and operators (up to 4 times more), without yielding improvement in effectiveness in most cases. They pointed out two groups of factors to explain most of the variations in the relevance of search results: users may fail using the appropriate term, or, they fail using the relevant operator when issuing their query.

In the context of digital humanities, our work is focused on digital libraries (DLs). When compared to web information retrieval (IR), IR in DLs has two specificities. (i) Corpora contents refer to domain specific topics (e.g., wildlife,

flora, art, sports). As a consequence, users' needs and vocabulary are usually more specific than in common corpora. (ii) Users are often accustomed to query operators (e.g., librarians, scholars) [1]. Let us consider the example of an expressive search *'Potato famine in Ireland after mid-19th century; prioritizing documents evoking the Connacht province; excluding those evoking Cork or any other location included in Cork city.'* The more expressive the query, the more criteria and corresponding operators are used. We consider the usual acceptance that Geographic Information gathers three dimensions, namely spatial, temporal and, topical [29]. Figure 1 illustrates these dimensions, as well as the variety of criteria within each dimension often used when searching for information in DLs. This is reinforced by Larson [14] when he demonstrates the effectiveness of Geographic Information Retrieval (GIR) methods in DLs.

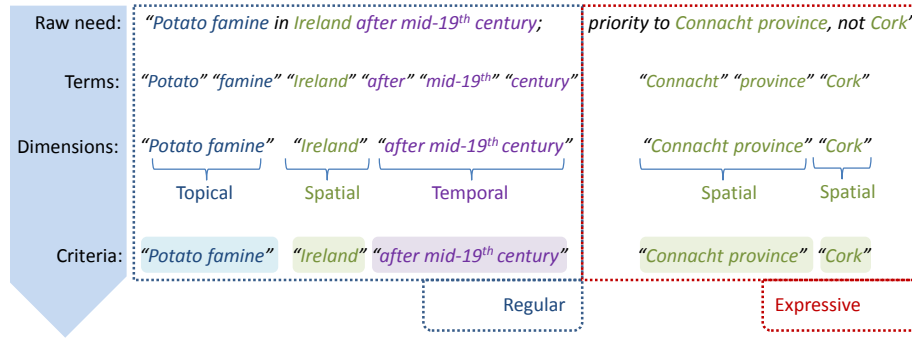


Fig. 1. Raw query as issued by the user, and segmentation into dimensions and criteria

In [21] we introduced a system supporting the basic part of such queries (regular part of the query in Fig. 1) with matching operators dedicated to each dimension. However, during the query formulation process, the user often introduces more expressiveness with modal operators such as *May*, *Should*, *Must*, *Must not*, to give priority to a criterion or to use a criterion as an exclusion key (expressive part of the query in Fig. 1). Therefore, in this paper, we propose the CMRP (Criterion, Matching, Requirement, Preference) multi-criteria IR framework to explore DLs. It supports the association of roles (*Exclusion*, *Demotion*, *Regular*, *Promotion*, *Necessity*, or *Target*) to any criterion in order to extend the expressiveness of queries (regular and expressive parts of the query in Fig. 1). We hypothesize that the association of roles to query criteria will improve result effectiveness. The CMRP framework is implemented within the geographic IR engine called PIV³, for 'Virtual Itineraries in the Pyrenees Mountains'⁴.

We developed spatial and temporal IR systems (IRSs) dedicated to unstructured textual documents [8,15]. Now, the challenge consists in aggregating spa-

⁴ The PIV project is funded by the Pau City Council and the MIDR multimedia library.

tial, temporal and topical IRSs within a single IR framework keeping in mind that such a framework may further integrate IRSs dedicated to documents of different formats. Moreover, the aggregation approach has to support expressive roles associated with any criterion.

The paper is organized as follows. In Sect. 2, we review the literature related to multi-criteria search and GIR systems, as they may benefit from our framework. In Sect. 3, we describe the CMRP framework and the PIV³ search engine: it presents the core components of PIV³—spatial, temporal, and topical IRSs—and their federation within PIV³. In Sect. 4, we report evaluations through experiments complying with the evaluation framework proposed in [21]. Finally, Sect. 5 concludes the paper and outlines research directions.

2 Related Work: Multi-Criteria Search and Geographic Information Retrieval

For classical IR, more expressive queries lead to better results. Indeed, Kelly and Fu [13] showed a strong relationship between query expansion (with additional information from users) and performance.

Croft and Harabagiu [4] promoted the combination of different text representations and search strategies as a technique for improving the effectiveness of information retrieval. They described three categories of approaches: the combination of representations given as input for retrieval algorithms, the combination of retrieval algorithms, and the combination of search system outputs. Fox and Shaw’s [7], as well as Fernandez et al.’s [6] score-based aggregation models may be classified in the third category. Both propose to normalize the source scores to a common distribution before being combined.

Therefore, we make the assumption that various search criteria plus advanced roles specifying whether a criterion is required to retrieve a document or even a criterion is used to promote, demote, or reject a document will improve descriptions of information needs and result effectiveness. In this section, we first focus on recent work dedicated to the aggregation of result lists issued from multi-criteria IR. Then, we briefly describe work related to multi-criteria IR dedicated to the handling of spatial and temporal dimensions of documents.

2.1 Multi-Criteria Information Retrieval

In the context of multi-criteria IR, Farah and Vanderpooten [5] define result aggregation as a process that ranks documents by combining document retrieval status values (RSVs), also known as ‘score,’ associated with each criterion. Figure 2 illustrates the principle of multi-criteria retrieval. A multi-criteria query conveys the user’s need. The search engine parses the corpus in order to match the query and document contents. A result list L_i comprising presumably relevant documents with corresponding RSVs is produced for each criterion. Then, the search engine aggregates those result lists to produce the final result L which is displayed to the user.

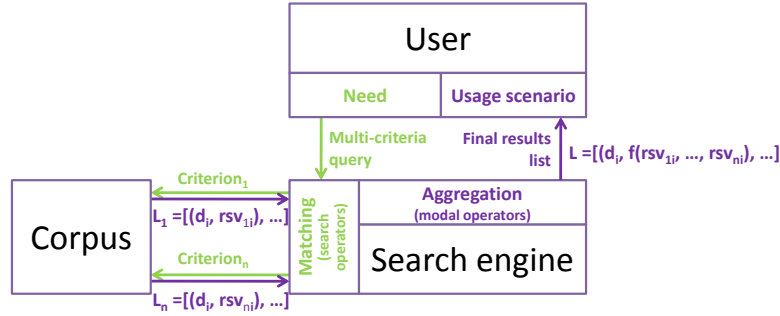


Fig. 2. Principle of multi-criteria retrieval

Farah and Vanderpooten [5] consider three classes of aggregation functions and describe them as follows:

Totally Compensatory Logic. This consists in computing a single RSV using functions such as the weighted sum. For such functions, a low score for one criterion can be compensated by one or several high scores for other criteria. So this kind of approach does not handle any required criterion. Prominent compensatory functions are Borda Count [2], CombMAX, CombMNZ, and CombSUM [7].

Non-Compensatory Logic. In this case, aggregation is mainly based on the value of one criterion, such as the lowest score of all criteria. The remaining criteria are only used to discriminate between documents with similar scores. A clear weakness of this class of operators is that a large part of the scores is ignored or plays a minor role. CombMIN [7] and Prioritized And [3] are functions of the non-compensatory class.

Partially Compensatory Logic. Farah and Vanderpooten [5] propose to combine any set of criteria while taking into consideration the imprecision underlying the query formulation process. In a similar way, da Costa Pereira et al. [3] propose the Prioritized Scoring Model (PSM) aggregation function giving the most important weight to the first criterion and decreasing weights to remaining criteria. Considering the important need of expressiveness for queries in the context of DLs, we make the assumption that partially compensatory functions supporting roles (e.g., exclusion, demotion, promotion, necessity) associated to query criteria may improve result effectiveness.

GIR is a domain where complex queries are issued, especially when searching for information in DLs [14]. Moreover, the ratio of geographic queries submitted to usual search engines is about 15% according to studies described in [25,9,12], and may increase in the context of DLs. Consequently, we propose a brief review of GIR frameworks in the next section.

2.2 Frameworks for Geographic Multi-Criteria IR

Among specialized fields of IR, Geographic IR (GIR) is concerned with multi-criteria queries. Indeed, a piece of geographic information is generally described as a set of three dimensions: spatial, temporal, and topical [29]. As a result, lots of GIR systems support multi-criteria geographic queries to handle those three dimensions.

Concerning non-compensatory aggregation, CITER [23] relies on a ‘parallel filtering’ approach: this system targets simultaneously and separately each dimension, and then combines the different result lists by computing their intersection. Because this parallel filtering operation is based on the Set theory, documents relevance is binary. As a result, the retrieved documents are not ranked.

Concerning compensatory aggregation, DIGMAP [19] relies on the ‘linear interpolation’ approach (e.g., arithmetic mean). Nevertheless, such aggregation models may be biased as they combine results returned by different IRSs characterized by specific information representation and score computation approaches.

To the best of our knowledge, all GIR systems use a non-compensatory or a compensatory aggregation; none use partially compensatory aggregation. For a more detailed presentation of current GIR systems we refer the reader to [21].

Systems tackling several dimensions for search purpose, like GIR systems, usually federate mono-dimensional IR systems, each one supporting its own domain-specific matching operators. Such multidimensional systems may benefit from partially compensatory aggregation functions supporting roles for refining the relative importance of criteria (Fig. 3). The next section presents the proposed CMRP framework, which provides a solution to design and process expressive queries.

3 CMRP: a Partially Compensatory Aggregation-Based IR Framework

CMRP is a GIR framework dedicated to textual corpora. It involves process flows dedicated to spatial and temporal information recognition and interpretation [8,15]. It builds several levels of index describing detailed spatial and temporal extracted entities, as well as larger points of view supported by spatial and temporal tile-based generalization approaches [22]. Though, it enables expressive queries, for which constraints can be attached to any criterion.

Let us consider again the example of the expressive search ‘*Potato famine in Ireland after mid-19th century; prioritizing documents evoking the Connacht province; excluding those evoking Cork or any other location included in Cork city.*’ Fig. 3 illustrates the variety of matching operators and roles. Matching operators are dimension-dependent (e.g., spatial, temporal, topical) whereas roles correspond to the expressiveness supported by the CMRP framework. Fig. 3 details dimensions and criteria of the query introduced in Fig. 1: predefined roles and/or personalized modal operators are associated with each criterion.



Fig. 3. Criterion roles and modal operators for an expressive query composition

As users have sophisticated needs, they may formulate expressive queries in two different ways. Regular users assign one of the various proposed roles (detailed in Sect. 3.1) to a criterion (e.g., *Must* is a mandatory role, *Must not* is an exclusion role, *May* is an enhancement one). Advanced users may directly combine the $+$, $-$, \wedge modal operators (Fig. 3) to refine these roles (e.g., the \wedge operator gradually penalizes or boosts a criterion – weighted between $[-1, 1]$ –). The next section introduces the CMRP framework.

3.1 Proposed CMRP Framework

We designed a framework to handle expressive multi-criteria queries. Let D denote the set of documents. This is based on the quadruplet (C, M, R, P) defined as follows:

- $C = (c_1, \dots, c_n)$, where c_i is a criterion expressed in the query.
- $M = (m_1, \dots, m_n)$, where $m_i : C \times D \rightarrow [0; 1]$. m_i refers to an external matching function used to compare any document $d \in D$ with criteria c_i . Various operators may be available in the CMRP framework, such as intersection (\cap), equality ($=$), inclusion (\subseteq), proximity (\sim). They correspond to the functions supported by the invoked IRSs.
- $R = (r_1, \dots, r_n)$ where requirement r_i may be mandatory ($+$), neutral, (N) or excluded ($-$).
- $P = (p_1, \dots, p_n)$ with $p_i \in \mathbb{R}$, where preference p_i allows a user to weight criterion c_i according to the corresponding level of importance he/she wants.

Table 1 illustrates the expressive query given in Fig. 3. In this example, notice that the temporal criterion is not mandatory, but boosted whenever it is encountered in a document. In addition, the topical criterion is mandatory and boosted. Notice that the preferences are chosen by the user.

We identified six predefined roles (Fig. 4) that may be associated with a criterion: exclusion (negative filtering), demotion, regular, promotion, necessity (positive filtering), and target. Table 2 and Fig. 4 illustrate these different scenarios that we detail in this section.

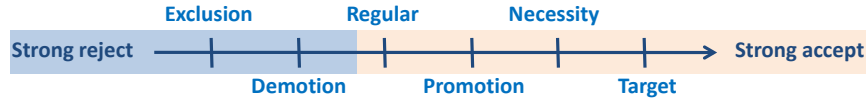
Table 1. Expressive query and result combination

| Quadruplet | Criterion | | | | |
|------------|---------------|---------|-----------|----------|-------------|
| | Topical | Spatial | Temporal | Spatial | Spatial |
| C | potato famine | Ireland | 1850-1900 | Connacht | Cork |
| M | = | \cap | \cap | = | \subseteq |
| R | + | + | N | N | - |
| P | 0.5 | 0 | 0.7 | 0.7 | 0 |

Table 2. Mapping of criteria roles to requirements and preferences

| Criteria roles | Requirements | | | Preferences | | |
|----------------|--------------|---|---|------------------|---|------------------|
| | - | N | + | \mathbb{R}_-^* | N | \mathbb{R}_+^* |
| Exclusion | ✓ | | | | ✓ | |
| Demotion | | ✓ | | ✓ | | |
| Regular | | ✓ | | | ✓ | |
| Promotion | | ✓ | | | | ✓ |
| Necessity | | | ✓ | | ✓ | |
| Target | | | ✓ | | | ✓ |

- A - Exclusion.** A user can decide to exclude one criterion. This criterion is used to reject documents and so it is not necessary to give it a preference.
- B - Demotion.** A user can choose to depreciate one criterion. Documents that contain this criterion are not dropped, but rather their global score are decreased.
- C - Regular.** A user can use a criterion without setting any preference or requirement.
- D - Promotion.** A user can decide to include an optional criterion. If this criterion is present, the document is more interesting, otherwise it is without consequences. Thus, a weight is associated with this criterion to influence the calculation of the global score.
- F - Necessity.** A user can decide to use a mandatory criterion. So, this criterion will not influence the global score calculation. This criterion must be present, but no preference is set.
- E - Target.** One criterion (or several) can be defined as the search goal. So, it must be present and it influences the global score calculation. If all criteria

**Fig. 4.** Six predefined possible roles refining the expressivity of a search criterion

are defined as targets with similar weights, they are all equally relevant for the search.

Note that some combinations are incompatible (for example, an exclusion $(-)$ requirement and a positive (\mathbb{R}_+^*) preference). As a result, it is mandatory to check the validity of each combination. Such combination scenarios are not allowed by the framework and are rejected.

The CMRP approach allows one to improve the expressiveness of each criterion in the query. The final score of a document is computed by a partially compensatory aggregation function (Eq. 1). This combines scores so that each criterion has an effect on the final score. Notice that, if a mandatory criterion is missing or if an excluded one is present in a document (denoted as $d_i \sim R$), this document is not retrieved (i.e., $RSV(d_i) = 0$). As a result, a mandatory or excluded criterion can not be compensated.

$$RSV(d_i) = \begin{cases} 0 & \text{if } d_i \sim R \\ \frac{1}{\sum_{i=1}^n p_i |p_i > 0} \cdot \sum_{i=1}^n p_i \cdot m_i(d_i, c_i) & \text{else} \end{cases} \quad (1)$$

To sum up, we introduced a new search process that extends those applied in criteria aggregation [5] in order to process multi-criteria queries. CMRP has several advantages:

- For each criterion, matching functions depend on dimension-dedicated IRSs.
- For each criterion, expressiveness (i.e., requirement or preference) is independent of the invoked IRS.
- For a query, result aggregation is partially compensatory: the equation for score computation takes into account the user's requirement and preference weights associated with each criterion.

Now we present this framework implementation within the PIV³ meta-engine that handles multi-criteria queries.

3.2 PIV³ Meta-Engine Implementing CMRP Framework

The CMRP framework can be implemented on top of a single search engine or with several search engines federated by a meta-search engine. Building on Rasolofo et al. [24], we followed the second approach to implement the CMRP framework within the PIV³ meta-search engine.

The PIV project aims to manage a digitized collection of documents published in the 19th century about the French Pyrenees Mountains. Following recent work on Digital Libraries [16], the main goal of the PIV project is to help users find accurate information inside books. It intends to overcome usual IRS limitations regarding geographic information management. Thus, we designed three process chains for spatial, temporal, and topical information indexing [8,15]. The generated indexes allow the retrieval of document units along with relevance scores consistent with any spatial, temporal, and topical dimension.

Figure 5 describes a meta-search engine as a broker, which splits a query into sub-queries dedicated to the targeted search engine. It also supports a result list aggregation process that produces a single list of results by merging several result lists.

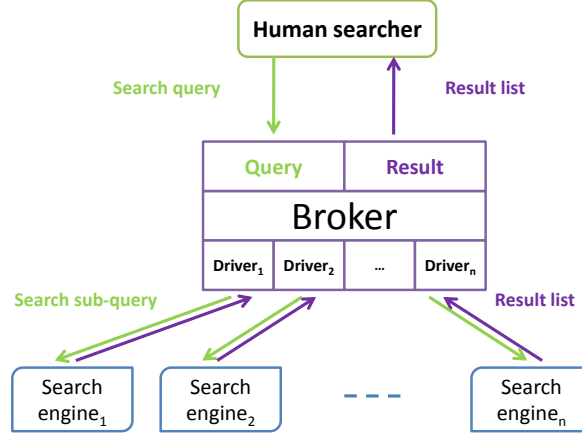


Fig. 5. A meta-search engine architecture according to Rasolofo et al. [24]

The PIV³ meta-search engine includes an instance of the CMRP framework (Sect. 3.1) within the global architecture described in Fig. 5. The PIV³ drivers feature three primary search engines: PIV_spatial, PIV_temporal, and PIV_topical (based on the Terrier [20] search engine). Spatial, temporal, and term-based specific matching functions are declared in these drivers. As a result, PIV³ federates three mono-dimensional IRSs. As it embeds dimension-specific operators, it splits a multi-criteria query into sub-queries and forwards them to the corresponding IRSs. Moreover, PIV³ supports the CMRP model to extend the expressiveness of the different criteria as defined in Sect. 3.1. It computes a partially compensatory aggregation to build a merged list of results comprising ranked documents.

Another important feature of this approach lies in the generalized information representations and retrieval models supported by PIV_spatial and PIV_temporal primary IRSs [21]. Each IRS supports a tile-based generalization of spatial or temporal information. Such an approach features a generalization resulting in an homogeneous and comparable representation of the various dimensions of geographic information. It is related to well-known IR models based on term lemmatization / truncation and weighting schemes (e.g., term frequency in the Vector Space Model). The final scores are not biased since representations and scores were generalized prior to the combination process. We refer the reader to [21] for further explanations on this subject.

In the next section, we evaluate the CMRP framework implemented in the PIV³ meta-search engine and compare it to CombMNZ and PSM aggregation functions.

4 Experiments

We aim at evaluating a search engine that handles expressive multi-criteria queries in a geographical context. The problem is, as explained in [21], that there is no test collection covering all geographic dimensions (spatial, temporal and topical). So, we propose to evaluate the CMRP framework, implemented in PIV³, according to two case studies:

1. **TREC test collection [30]**. Multi-criteria queries are submitted to Terrier search engine (with TF-IDF IR model) whereas operators are managed by PIV³. Thus, the CMRP framework can be used to extend topical search engines that do not support advanced operators or that support only some of them;
2. **MIDR.2010 test collection [21]**. Multi-criteria queries are submitted to PIV_Spatial, PIV_Temporal and Terrier search engines that are federated within PIV³.

During these experiments, queries were built as follows:

- Criteria were listed explicitly by the user;
- Then, expressiveness was specified upon each criterion (matching operators / modal roles) by an automated topic variant generator (case study 1) or by the user (case study 2).

4.1 Topical multi-criteria IR

As we did in [11], we conducted experiments on the TREC-8 [30] standard test collection. The TREC-8 corpus of newspaper articles corresponds to the kind of document units (e.g., newspaper or travelogue paragraphs) that the PIV³ engine is expected to retrieve. Moreover, it provides $n = 50$ topics covering various subjects, allowing us to realize significance testing. Each topic is worded as:

- A *title*: a sequence of two or three terms (in general) that would be submitted as a query to a search engine by an average user.
- A *description*: a few sentences describing the user’s information needs in plain text.
- A *narrative*: a longer text than the description, which gives insights into the intent of the user, and unambiguously states what information is relevant or irrelevant for the searcher.

We demonstrated the benefits of query operators (+ and ^) on regular topics in [11]. Here, because topics contain few terms and none that can be associated with exclusion or demotion roles, we propose to use the narrative part to extend each topic title with new terms. These terms are split into two categories:

1. positive terms are associated randomly with necessity (+), promotion ($\textcircled{+}$) and target roles;
2. negative terms are associated randomly with exclusion (-) and demotion ($\textcircled{-}$) roles.

Therefore, the experiment evaluates the three following scenarios:

1. *Regular topic*. We consider this as the baseline using the original titles of the TREC-8 test collection.
2. *Extended topic*. This is comprised of the positive terms only from title and some manually extracted from the narrative. No operator is associated with them.
3. *Expressive topic*. This is composed of the terms composing both positive and negative categories. Operators are randomly associated with these terms.

Within the *Expressive topic* scenario, we intend to check that any topic can be rewritten with operators, such that it leads to more accurate search results. Then, as explained in [11], a **Topic Variant Generator** is used to generate up to 6,561 variants (i.e., 4 distinct terms), per topic, due to all the possible combinations of the terms with potential roles. For the 50 topics we have 188,973 variants. The following example (Table 3) illustrates some variants generated for the *Hurricanes but not damages* topic. Generating variants for t_p positive terms leads to 9^{t_p} possibilities ($+t_{p1}$, $t_{p1}^{\textcircled{+0.25}}$, $t_{p1}^{\textcircled{+0.5}}$, $t_{p1}^{\textcircled{+0.75}}$, $t_{p1}^{\textcircled{+1}}$, $+t_{p1}^{\textcircled{+0.25}}$, $+t_{p1}^{\textcircled{+0.5}}$, $+t_{p1}^{\textcircled{+0.75}}$, $+t_{p1}^{\textcircled{+1}}$) and for t_n negative terms leads to 6^{t_n} possibilities (\emptyset , $-t_{n1}$, $t_{n1}^{\textcircled{-0.25}}$, $t_{n1}^{\textcircled{-0.5}}$, $t_{n1}^{\textcircled{-0.75}}$, $t_{n1}^{\textcircled{-1}}$). Therefore, the total of variants is $9^{t_p} \times 6^{t_n}$.

Table 3. Excerpt of a topic variants generated with random combinations: positive terms associated with necessity, promotion and target roles and negative terms associated with exclusion and demotion roles

| Variant # | Query variants | |
|-----------|---|--|
| 1 | +hurricane | \emptyset |
| 2 | +hurricane | -damage |
| 3 | +hurricane | damage $^{\textcircled{-0.25}}$ |
| 4 | +hurricane | damage $^{\textcircled{-0.5}}$ |
| \vdots | \vdots | \vdots |
| 54 | +hurricane $^{\textcircled{+1}}$ | damage $^{\textcircled{-1}}$ |

We queried PIV³ with each of the *regular*, *expressive* and *extended topics* corresponding to the 50 TREC-8 topics. To study the distribution of effectiveness data values we use ‘boxplot’ visualization [28,32] with the Gnuplot software⁵. A boxplot (a.k.a. box-and-whisker diagram) summarizes several descriptive statistics. The interquartile range (IQR) spans the lower quartile to the upper quartile.

⁵ <http://www.gnuplot.info/>

The middle 50% of the ranked data lies in the IQR. It is represented as a *box* (central rectangle), which shows the spread of data values. The median is shown as a *segment* inside the box. This is the middle half of the data values, and allows one to assess the symmetry of the distribution. The *whiskers* extend from the ends of the box to the most distant value lying within $1.5 \times \text{IQR}$. Larger and lower values are considered as outliers; these are plotted with *black circles*.

Figure 6 presents effectiveness results for the three groups of topics. These results can be organized in 4 categories:

- Expressive topics improve upon regular topics but extended topics are quite similar (less than 5% of difference between results of expressive and extended topics) for 6% of all cases. Regular topic expansion with new terms is enough (e.g., queries number 2, 8, and 9; Fig. 6).
- Expressive topics overcome extended topics, as well as regular ones for 26% of all cases. Several (one or more) of the randomly produced term/operator combinations improve retrieval results (e.g., queries number 33, 38, and 39; Fig. 6).
- Expressive topics improve upon regular topics, when extended topics do not overcome regular ones, for 32% of all cases. Expansion based on new terms does not improve the baseline. However, when associated with convenient operators, these terms improve retrieval results (e.g., queries number 10, 29, and 34; Fig. 6).
- Expressive topics overcome extended topics but not regular ones for 32% of all cases. Expansion based on new term addition is not convenient although operators improve significantly their retrieval results (e.g., queries number 6, 30, and 31; Fig. 6).

Table 4. Experiment comparing different scenarios

| Topics | MAP | Improvement (%) | <i>t</i> -test (<i>p</i> -value) |
|--------------------|--------|--------------------|--------------------------------------|
| Regular (baseline) | 0.1970 | 0.00 | 0.0000 |
| Extended | 0.1367 | −30.61 | 0.0097 |
| Expressive | 0.2502 | 26.96 | 0.0098 |

As shown in Tab. 4, Extended topics reduce results precision (−30.61%), whereas expressive topics improve results (+26.63%), and it is statistically significant (*t*-test with $p < 0.05$). If we compare expressive topics and extended topics (considering that expressive topics add roles to extended ones), the improvement is even bigger (+82.98%).

Several reasons can explain why extended topics are worse than the baseline: (1) the original query is already ideal so extension only brings noise; (2) the original query can be improved, but adding keywords with similar importance (totally compensatory approach) only brings noise. Nevertheless, these keywords associated with modal operators bring suited expressiveness and improve results. It shows the difficulty to choose the good terms.

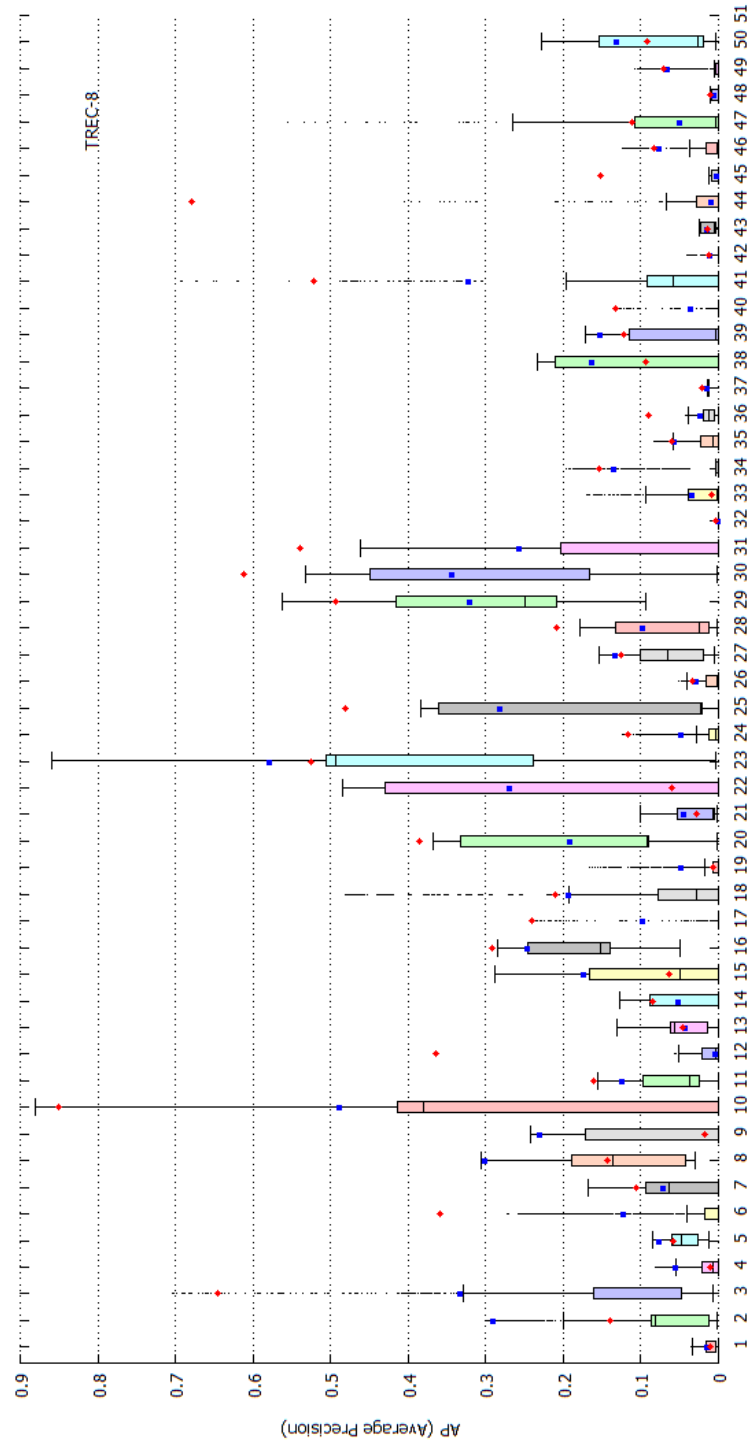


Fig. 6. Average Precision (AP) boxplots showing potential for improvement with expressive topics on 50 queries (TREC-8). Diamonds ◆ show the AP of regular topics, diamonds ■ show the AP of extended topics and the higher positions of boxplots indicate the AP corresponding to the best topic variant

Overall, notice that, among all the variants of an expressive topic, the best one implies different promotion roles (92 % of all cases) and a necessity role (80% of all cases). Moreover, 50% of topics involve exclusion or demotion criteria and about 40% of them contribute to topic improvement.

Therefore, when terms of an expressive topic are well chosen there always exists a multi-criteria query that may improve the retrieval results. As geographical queries often involve two or three dimensions, we believe they are more suited to such a multi-criteria querying approach: specific spatial and temporal search criteria are easier to identify than term-based criteria and corresponding roles are also easier to specify. The next experiment evaluates PIV³ engine with a geographical test collection. So we verify that with a better suited context, PIV³ will improve effectiveness in a larger extent.

4.2 Geographic Multi-criteria IR

We used the MIDR.2010 cultural heritage test collection [21] (available on the PIV project website⁶), whose corpus represents 5,645 paragraphs extracted from 11 books published between the 18th and 20th centuries. We manually designed 10 expressive queries targeting geographic items present in this DL (e.g., Fig. 1 query: ‘*Potato famine in Ireland after mid-19th century; priority to Connacht province, not Cork*’). Moreover, we measured effectiveness of the tested IRSs with respect to *Mean Average Precision* (MAP) [18]. Let us note that in [21] we only experimented the CombMNZ totally compensatory aggregation function.

Table 5. First experiment comparing aggregation approaches

| Model | MAP | Improvement (%) | <i>t</i> -test (<i>p</i> -value) |
|-------------------------------------|--------|--------------------|--------------------------------------|
| CombMNZ (baseline) [7] | 0.1658 | 0.00 | 0.00 |
| Prioritized Scoring Model (PSM) [3] | 0.1034 | −37.64 | 0.18 |

Here, we first compare a full compensatory aggregation function (CombMNZ) [7] to the partially compensatory aggregation function, ‘Prioritized Scoring Model’ (PSM), proposed by da Costa Pereira et al. [3](Sect. 2.1). The main advantage of the PSM function is that a user just needs to specify the order of criteria since decreasing weights of each criterion are calculated automatically. Although PSM is intended to improve personalized search, it is also appropriate for dealing with expressive queries. It is thus interesting to compare its effectiveness to the CombMNZ function and to the CMRP model designed to enhance search with expressive queries. With the PSM approach, the query of Fig. 1 can be expressed as: ‘potato famine > Ireland > 1850-1900 > Connacht’. ‘not Cork’ cannot be expressed because exclusion is not supported in PSM. Nevertheless, in this first experiment, as shown in Tab. 5, PSM yields weaker results than the

⁶ <http://t2i.univ-pau.fr/MIDR/>

CombMNZ baseline. This difference is not statistically significant however (t -test with $p > 0.05$). PSM is a partially compensatory aggregation function however it does not support expressive queries involving advanced roles like necessity, target, exclusion, demotion. Therefore, in such a context, a full compensatory aggregation function returns better results.

Second, we compare the CMRP framework based on a partially compensatory aggregation and implemented by PIV³ to the CombMNZ baseline with the same queries. As shown in Tab. 6, the CMRP framework yields an improved effectiveness by 54.16 %, which is not statistically significant however ($p > 0.05$). We need to continue our experiment (more than 10 queries) to have statistically significant results. The observed difference may be due to the fact that CMRP allows users to refine roles for each criterion, unlike CombMNZ. Indeed, CombMNZ does not support requirement or preference modal operators. As the CMRP framework allows users to refine several roles, it enables enhanced expressiveness for queries.

Table 6. Second experiment comparing aggregation approaches

| Model | MAP | Improvement (%) | t -test (p -value) |
|--------------------------|--------|--------------------|----------------------------|
| CombMNZ (baseline) [7] | 0.1658 | 0.00 | 0.00 |
| CMRP (PIV ³) | 0.2556 | +54.16 | 0.10 |

Both experiment settings rely on documents retrieved with quite low average precisions (MAP). . This is due to several restrictive topics in the test collection. However, systems participating in TREC have similar results [30].

To sum up, the first experiment of this section showed that it is better to use a full compensatory aggregation function (e.g., CombMNZ) than to use a function that does not take into account the expected expressiveness. The second experiment showed that CMRP framework yields better results (+54.16 %) as it provides the user with a way to express advanced roles for each criterion.

5 Conclusion and Future Work

In this paper, we considered expressive user needs formulated in various contexts, such as DLs. Users spend efforts and time creating expressive queries. This process leads to multi-criteria queries, where each criterion may be refined by a role describing its importance and semantics within the query. We proposed the CMRP framework for handling preferences and requirements in search queries. Requirements are achieved by mandatory, exclusion, or neutral modal operators whereas preferences are achieved by positive or negative amounts associated to a boosting modal operator in order to promote or demote a document. CMRP integrates them in a partially compensatory aggregation function computed to

retrieve a single result list from different criteria-based result lists. As a proof of concept, we designed the PIV³ GIR meta-search engine, which implements the CMRP framework. This system takes spatial, temporal, and topical dimensions into account, so as to provide more expressiveness to users who express their information needs into queries. Each domain-specific IRS supports spatial, temporal, or topical matching functions whereas the meta-search engine allows the end-user to associate any criteria with expressive roles. A first experiment based on TREC-8 topical queries shows that PIV³ significantly improves search results (+26.96 %). A second experiment based on geographical queries, and the cultural heritage test collection MIDR_2010 shows that PIV³ GIRS outperforms (+54.16 %) CombMNZ [7].

These corpora and dedicated IR systems deal with unstructured textual documents. However, our framework may integrate other IRSs dedicated to geographic information contained in documents of different formats: image, video or sound. PIV³ architecture may federate such various IR systems; the only constraint is to build the corresponding drivers.

Extending the number and the scope of IR matching operators, as well as making result aggregation formula customizable, makes it difficult for end-users to control the scope of such queries. We plan to investigate processes supporting user needs description in a textual mode. A first stage may call the federated IR systems to extract criteria automatically (e.g. spatial and temporal entities and keywords) from a textual query. A second stage should ask the user to validate/adapt each criterion. Finally, a third stage may guide the user to specify expressiveness attached to each criterion. Indeed, our PIV_spatial and PIV_temporal IRSs process textual queries to extract spatial and temporal entities automatically [8,15].

We plan to propose two approaches for the expressiveness specification stage. A first one, dedicated to regular users, would allow them to choose among predefined roles in order to add expressiveness to any criterion. A second one, dedicated to advanced users, may propose them to customize explicitly their criteria with modal (\wedge , $+$, $-$) and matching ($=$, \subseteq , \cap) operators.

Regarding the MIDR_2010 test collection, a larger corpus and query collection need to be constituted for conducting further experiments.

Moreover, one can note that the proposed language completeness is not ensured although CMRP supports “or” and “and” operators via requirement and preference operators. However, this evaluation delivers some interesting preliminary findings and this contribution may stimulate the use of unexploited operators in GIRSs, as well as in other IRSs (e.g., Lucene [10]).

References

1. Aula, A., Khan, R.M., Guan, Z.: How does search behavior change as search becomes more difficult? In: CHI’10: Proceedings of the 28th international conference on Human factors in computing systems. pp. 35–44. ACM, New York, NY, USA (2010)

2. de Borda, J.C.: Mémoire sur les élections au scrutin. In: *Histoire de l'Académie Royale des Sciences*. Paris (1781)
3. da Costa Pereira, C., Dragoni, M., Pasi, G.: Multidimensional relevance: A new aggregation criterion. In: *ECIR'09: Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*. pp. 264–275. Springer-Verlag, Berlin, Heidelberg (2009)
4. Croft, B., Harabagiu, S.: *Advances in Information Retrieval: Recent Research from the Center for Intelligent Information Retrieval*. Kluwer Academic Publishers (2000)
5. Farah, M., Vanderpooten, D.: An outranking approach for information retrieval. *Inf. Retr.* 11(4), 315–334 (2008)
6. Fernández, M., Vallet, D., Castells, P.: Probabilistic score normalization for rank aggregation. In: *ECIR'06*. pp. 553–556 (2006)
7. Fox, E.A., Shaw, J.A.: Combination of Multiple Searches. In: Harman, D.K. (ed.) *TREC-1: Proceedings of the First Text REtrieval Conference*. pp. 243–252. NIST, Gaithersburg, MD, USA (Feb 1993)
8. Gaio, M., Sallaberry, C., Etcheverry, P., Marquesuzaa, C., Lesbegueries, J.: A global process to access documents' contents from a geographical point of view. *J. Vis. Lang. Comput.* 19(1), 3–23 (2008)
9. Gan, Q., Attenberg, J., Markowetz, A., Suel, T.: Analysis of geographic queries in a search engine log. In: *LocWeb'08: Proceedings of the first international workshop on Location and the web*. pp. 49–56. ACM, New York, NY, USA (2008)
10. Gospodnetić, O., Hatcher, E.: *Lucene in Action*. Manning Publications (2005)
11. Hubert, G., Cabanac, G., Sallaberry, C., Palacio, D.: Query operators shown beneficial for improving search results. In: Gradmann, S., Borri, F., Meghini, C., Schuldt, H. (eds.) *TPDL. Lecture Notes in Computer Science*, vol. 6966, pp. 118–129. Springer (2011)
12. Jones, R., Zhang, W.V., Rey, B., Jhala, P., Stipp, E.: Geographic intention and modification in web search. *Int. J. Geogr. Inf. Sci.* 22(3), 229–246 (2008)
13. Kelly, D., Fu, X.: Eliciting better information need descriptions from users of information search systems. *Inf. Process. Manage.* 43(1), 30–46 (2007)
14. Larson, R.R.: Geographic Information Retrieval and Digital Libraries. In: *ECDL'09: Proceedings of the 13th European Conference on Digital Libraries*. LNCS, vol. 5714, pp. 461–464. Springer (2009)
15. Le Parc-Lacayrelle, A., Gaio, M., Sallaberry, C.: La composante temps dans l'information géographique textuelle. *Document Numérique* 10(2), 129–148 (2007)
16. Liesaputra, V., Witten, I.H., Bainbridge, D.: Searching in a Book. In: *ECDL'09: Proceedings of the 13th European Conference on Digital Libraries*. LNCS, vol. 5714, pp. 442–446. Springer (2009)
17. Lucas, W., Topi, H.: Form and function: the impact of query term and operator usage on Web search results. *J. Am. Soc. Inf. Sci. Technol.* 53(2), 95–108 (Jan 2002)
18. Manning, C.D., Raghavan, P., Schütze, H.: *Introduction to Information Retrieval*. Cambridge University Press (Jul 2008)
19. Martins, B., Borbinha, J., Pedrosa, G., Gil, J.a., Freire, N.: Geographically-aware information retrieval for collections of digitized historical maps. In: *GIR'07: Proceedings of the 4th ACM workshop on Geographical information retrieval*. pp. 39–42. ACM, New York, NY, USA (2007)
20. Ounis, I., Amati, G., Plachouras, V., He, B., Macdonald, C., Johnson, D.: Terrier Information Retrieval Platform. In: *ECIR'05: Proceedings of the 27th European Conference on IR Research*. LNCS, vol. 3408, pp. 517–519. Springer (2005)

21. Palacio, D., Cabanac, G., Sallaberry, C., Hubert, G.: On the evaluation of geographic information retrieval systems: Evaluation framework and case study. *International Journal of Digital Libraries (IJDL)* pp. 91–109 (2010)
22. Palacio, D., Sallaberry, C., Gaio, M.: Generalize geographic information to combine ir results. In: *AGILE: The 14th AGILE International Conference on Geographic Information Science – Poster session*. pp. 151–153. Utrecht, The Netherlands (2011)
23. Pfoser, D., Efentakis, A., Hadzilacos, T., Karagiorgou, S., Vasiliou, G.: Providing Universal Access to History Textbooks: A Modified GIS Case. In: *W2GIS'09: Proceedings of the 9th International Symposium on Web and Wireless Geographical Information Systems*. LNCS, vol. 5886, pp. 87–102 (2009)
24. Rasolofo, Y., Hawking, D., Savoy, J.: Result merging strategies for a current news metasearcher. *Inf. Process. Manage.* 39(4), 581–609 (Jun 2003)
25. Sanderson, M., Kohler, J.: Analyzing Geographic Queries. In: *SIGIR-GIR'04: Proceedings of the Workshop on Geographic Information Retrieval at SIGIR* (2004)
26. Silverstein, C., Marais, H., Henzinger, M., Moricz, M.: Analysis of a very large web search engine query log. *SIGIR Forum* 33(1), 6–12 (Sep 1999)
27. Spink, A., Wolfram, D., Jansen, M.B.J., Saracevic, T.: Searching the web: the public and their queries. *J. Am. Soc. Inf. Sci. Technol.* 52(3), 226–234 (Feb 2001)
28. Tukey, J.W.: *Exploratory data analysis*. Addison-Wesley (1977)
29. Usery, E.L.: A feature-based geographic information system model. *Photogramm. Eng. Rem. Sens.* 62(7), 833–838 (1996)
30. Voorhees, E.M., Harman, D.K.: Overview of the Seventh Text REtrieval Conference (TREC-8). In: *TREC-8: Proceedings of the 8th Text REtrieval Conference*. pp. 1–23 (1999)
31. White, R.W., Morris, D.: Investigating the querying and browsing behavior of advanced search engine users. In: *SIGIR'07: Proceedings of the 30th annual international ACM SIGIR conference*. pp. 255–262. ACM, New York, NY, USA (2007)
32. Williamson, D.F., Parker, R.A., Kendrick, J.S.: The box plot: A simple visual method to interpret data. *Ann. Intern. Med.* 110(11), 916–921 (Jun 1989)